Splicing of a divergent subclass of AT-AC introns requires the major spliceosomal snRNAs.

Q Wu and A R Krainer

RNA 1997 3: 586-601

RNA (1997), 3:586–601. Cambridge University Press. Printed in the USA.

Splicing of a divergent subclass of AT-AC introns requires the major spliceosomal snRNAs

QIANG WU and ADRIAN R. KRAINER

Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, USA,
and the Program in Molecular and Cellular Biology, State University of New York at Stony Brook,
Stony Brook, New York 11794, USA

ABSTRACT

AT-AC introns constitute a minor class of eukaryotic pre-mRNA introns, characterized by 5′-AT and AC-3′ boundaries, in contrast to the 5′-GT and AG-3′ boundaries of the much more prevalent conventional introns. In addition to the AT-AC borders, most known AT-AC introns have highly conserved 5′ splice site and branch site sequence elements of 7–8 nt. Intron 6 of the nucleolar P120 gene and intron 2 of the SCN4A voltage-gated skeletal muscle sodium channel are AT-AC introns that have been shown recently to be processed via a unique splicing pathway involving several minor U snRNAs. Interestingly, intron 21 of the same SCN4A gene and the corresponding intron 25 of the SCN5A cardiac muscle sodium channel gene also have 5′-AT and AC-3′ boundaries, but they have divergent 5′ splice site and presumptive branch site sequences. Here, we report the accurate in vitro processing of these two divergent AT-AC introns and show that they belong to a functionally distinct subclass of AT-AC introns. Splicing of these introns does not require U12, U4atac, and U6atac snRNAs, but instead requires the major spliceosomal snRNAs U1, U2, U4, U5, and U6. Previous studies showed that G → A mutation at the first position and G → C mutation at the last position of a conventional yeast or mammalian GT-AG intron suppress each other in vivo, suggesting that the first and last bases participate in an essential non-Watson–Crick interaction. Our results show that such introns, hereafter termed AT-AC II introns, occur naturally and are spliced by a mechanism distinct from that responsible for processing of the apparently more common AT-AC I introns.

Keywords: divergent AT-AC introns; in vitro splicing; pre-mRNA splicing; skeletal and cardiac muscle sodium channel

INTRODUCTION

Splicing takes place in a large ribonucleoprotein complex known as the spliceosome. Spliceosome assembly is a dynamic process that involves the ordered binding of several snRNAs and numerous proteins to the pre-mRNA substrate (Moore et al., 1993; Black, 1995; Reed, 1996). Splicing is thought to be catalyzed by snRNAs (Madhani & Guthrie, 1994; Newman, 1994; Nilsen, 1994) with the help of protein splicing factors (Fu, 1995; Krämer, 1996; Manley & Tacke, 1996). First, with the help of SR proteins and U2AF, the U1 and U2 snRNAs base pair with the conserved 5′ splice site and branch site, respectively (Kohtz et al., 1994; Valcarcel et al., 1996). Then U4, U5, and U6 snRNAs join the spliceosome as a tri-snRNP particle; U6 snRNA replaces U1 snRNA by base pairing to the 5′ splice site, which serves as a proof-reading step for 5′ splice site selection, and U1 and U4 are released from the spliceosome (Ares & Weiser, 1995). Finally, U5 snRNA aligns the two exons together through weak base pairing between the uridine-rich loop of U5 and nonconserved exon sequences (O'Keefe et al., 1996). As a result of two consecutive trans-esterification reactions, the intron is excised out as a lariat molecule, and the exons are joined together.

The great majority of introns have canonical 5′ splice sites that begin with GT and 3′ splice sites that end with AG. A minor class of introns with noncanonical AT 5′ splice sites and AC 3′ splice sites has also been found in several eukaryotic genes (Jackson, 1991; Hall & Padgett, 1994). These can be subdivided into those that have 5′ splice site and branch site elements with a close match to the respective consensus sequences, and those that have more divergent sequences. We refer to the former as AT-AC I introns, and to the latter as AT-AC II introns. A compilation of the AT-AC I introns known to date is shown in Table 1. They include 17 sequences that belong to 13 unrelated gene families;

Reprint requests to: Adrian R. Krainer, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, USA; e-mail: krainer@cshl.org.

TABLE 1. Compilation of known AT-AC I introns and their conserved sequences.[a]

Species	Gene	Intron number	5' splice site	Presumptive branch site	3' splice site	Size (nt)
Human	P120	6	GG atatcctt	...gttccttaac	aggcccac AT	99
Human	CMP	7	AC atatccct	...tctccttaac	tctgagtccac TG	615
Human	SCN4A	2	GC atatcctg	...ttccttgac	cctgccccac GC	126
Human	SCN5A	3	TC atatcc.....	?	cccacgcac GC	?
Human	CACNL1A4	1	CC atatcctt	...ttccttaat	tccccaatac TC	?
Human	GT335	6	TC atatcctt	...ttccgtaac	tctgcacac GA	3000
Human	E2F1	4	CG atatcctt	...ggtccttgac	tctgccaac CC	660
Mouse	E2F4	3	TG atatcctc	...tctccttaac	ccccacac CT	299
Mouse	SCN8A	2	TC atatcctt	...gccccttaac	tcctctctac AG	?
Mouse	CDK5	9	CG atatcctc...	?	acatggacac AC	440
Mouse	REP-3	6	AG atatcctt	...ttcttaat	cattactac AT	1400
Mouse	XPG	13	GA atatcctt	...aaaccttaac	tgccagcac AT	900
Mouse	HPS	15	TC atatcctc	...gccccttaac	caggcctgcac CT	600
Xenopus	TFIIS.oA	6	AG atatcctt	...attccttaaa	tccctttgcac CG	600
Fly	Prospero	2	CT atatcctt	...aatccttgac	tcctttgcac TC	87
Arabid.	ArLIM15	14	CC atatcctt	...ttccttaactgaagaaatcac AG		99
Arabid.	AtG5	7	AA atatcctt	.....atattaac	caaggcttaa GT	138
AT-AC I consensus			atatcctY	tccttraY	Yac	
GT-AG consensus		AG	gtaagt	ynytray	Y_{11}nyag	GT

[a]Sequences around the 5′ splice site, 3′ splice site, and presumptive branch site of known AT-AC I introns are shown. Smaller subsets of these sequences were shown in previous compilations (Jackson, 1991; Hall & Padgett, 1994; Wu & Krainer, 1996). P120: proliferating cell nucleolar antigen; AT-AC I intron also present in monkey, mouse, and dog homologues (Larson et al., 1990; Hall & Padgett, 1994). CMP: cartilage-matrix protein (Jenkins et al., 1990; size determined from our unpublished sequencing data); AT-AC I intron also conserved in chicken homologue (Kiss et al., 1989); SCN4A: skeletal muscle voltage-gated sodium channel α subunit; AT-AC I intron also conserved in rat homologue (McClatchey et al., 1992; George et al., 1993); SCN5A: cardiac muscle voltage-gated sodium channel α subunit (Wang et al., 1996); SCN8A: brain and spinal cord voltage-gated sodium channel α subunit (Kohrman et al., 1996); CACNL1A4: voltage-gated calcium channel α subunit (Ophoff et al., 1996); GT335: gene of unknown function (Lafrenière et al., 1996); E2F: family of transcription factors involved in transcription initiation of cell-cycle regulated genes (Neuman et al., 1996; R. Rempel, pers. comm.); CDK5: a cyclin-dependent kinase (Ohshima et al., 1995); REP-3: a protein thought to be involved in DNA repair (Liu et al., 1994); AT-AC I intron also present in human homologue, designated MSH3, but ends with AA instead of AC (Watanabe et al., 1996); XPG: DNA excision repair gene, defective in *Xeroderma pigmentosum* (Ludwig et al., 1996); HPS: mouse pale ear gene, homologous to human gene defective in Hermansky–Pudlak syndrome; AT-AC I intron also conserved in human HPS gene (Bailin et al., 1997; Feng et al., 1997); TFIIS.oA: transcription elongation factor; AT-AC I intron also present in closely related TFIIS.oB allele (Plant et al., 1996); Prospero: a homeodomain transcription factor expressed in neuronal precursor cells (Hall & Padgett, 1994); ArLIM15: *Arabidopsis thaliana* gene homologous to *Lilium longiflorum* LIM15 and to other meiosis-specific recombination genes related to *E. coli recA* (Sato et al., 1995; Wu et al., 1996); AtG5: *A. thaliana* putative transmembrane protein (Wu et al., 1996). Intron sequences are shown in lower case, flanked by two exon nucleotides on each side, in upper case. Invariant or virtually invariant sequences are shown as white letters on black background; less conserved sequences are shown as black letters on grey background. The underlined adenosine in the P120 sequence denotes the only AT-AC I branchpoint mapped to date (Tarn & Steitz, 1996a). Question marks indicate unknown sequences; dots indicate sequences that are not known or not shown. Some of the intron sizes are only approximate. The consensus sequences are shown at the bottom, with the conventional GT-AG intron consensus sequences also shown for comparison; r = purine; y = pyrimidine.

the AT-AC I introns are conserved in several species homologues that are not shown in the table. AT-AC I intron processing has been shown to require the minor U12 snRNA and probably also the minor U11 snRNA and the major U5 snRNA, but not the major U2, U4, or U6 snRNAs (Hall & Padgett, 1996; Tarn & Steitz, 1996a; Wu & Krainer, 1996; reviewed in Kreivi & Lamond, 1996). Two novel minor snRNAs, U4atac and U6atac, were recently identified in an AT-AC I spliceosome assembled in vitro with P120 pre-mRNA and shown to function in a manner analogous to that of U4 and U6 in the conventional spliceosome (Tarn & Steitz, 1996b; reviewed in Nilsen, 1996). Although the basal splicing of a sodium channel AT-AC I intron does not require U1 snRNA in vitro, a strong enhancement of AT-AC I splicing can result from exon-definition interactions between the AT-AC I spliceosome and U1 snRNP bound at a downstream, conventional 5′ splice site (Wu & Krainer, 1996).

An AT-AC I intron is present in several members of the voltage-gated sodium and calcium channel family (Table 1). Intron 2 of human SCN4A, the skeletal mus-

cle sodium channel α subunit (McClatchey et al., 1992; George et al., 1993) and intron 2 of its rat homologue (George et al., 1993), intron 2 of mouse SCN8A, the brain and spinal cord sodium channel α subunit (Kohrman et al., 1996), intron 3 of human SCN5A, the cardiac muscle sodium channel α subunit (Wang et al., 1996), and intron 1 of human CACNL1A4, a calcium channel α subunit (Ophoff et al., 1996), all belong to the AT-AC I subclass. In addition to the AT-AC I introns listed here, the published 5′ splice site sequence of intron 2 of human CACNL1A1, the fibroblast voltage-gated L-type calcium channel α subunit, precisely matches the AT-AC I consensus (Soldatov, 1994), whereas the 5′ splice site of intron 1 in two other calcium channel α subunits, CACNL1A2 (Yamada et al., 1995) and CACNL1A3 (Hogan et al., 1996), is reported to begin with GTATCCTT (rather than ATATCCTT); however, all three calcium channel introns reportedly end with the conventional AG 3′ splice site. The sodium channel family and calcium channel family genes are derived from a common ancestral gene and belong to the voltage-gated ion channel superfamily (Koester, 1991; Goldin, 1995); they have considerable nucleotide and amino acid sequence homology, and the unusual intron interrupts a homologous position of the coding sequence in all seven genes; moreover, all seven introns have the AT-AC I presumptive branch site consensus element (TCCTTRAC). Although the precise nucleotide sequence at the 5′ and 3′ splice sites should be re-confirmed in light of the discovery of bona fide AT-AC introns, we suspect that these unusual introns may be AT-AC I introns, and predict that similar AT-AC I introns will also be found at the corresponding position of other voltage-gated sodium and calcium channel α-subunit genes.

Previously, we proposed that intron 25 of the human SCN5A gene and intron 21 of the human SCN4A gene might belong to a distinct subclass of AT-AC introns (Wu & Krainer, 1996). This hypothesis was based on the fact that, although each of these introns begins with 5′-AT, the rest of the 5′ splice site does not match the highly conserved AT-AC I consensus sequence, ATATCCTY (George et al., 1993; Wang et al., 1996). Moreover, intron 21 of SCN4A lacks the highly conserved branch site sequence present in other AT-AC I introns, TCCTTRAC, in which the underlined A is presumed to be the branch nucleotide, as was shown to be the case for the P120 AT-AC I intron (Tarn & Steitz, 1996a). Likewise, intron 25 of the SCN5A gene lacks a conserved AT-AC I branch consensus (Table 2 and data not shown). These two introns interrupt a homologous position of the sequence of two members of the voltage-gated sodium channel family, and hence similar AT-AC II introns may be present in other members of the sodium channel gene family. Interestingly, an AT-AC II intron and an AT-AC I intron coexist with multiple conventional GT-AG introns in the SCN4A and SCN5A sodium channel genes. Mutations in the cardiac SCN5A gene are responsible for the cardiovascular long-QT syndrome (Wang et al., 1995), whereas mutations in the skeletal muscle SCN4A gene cause hyperkalemic periodic paralysis and paramyotonia con-

TABLE 2. Compilation of known AT-AC II introns.[a]

Species	Gene	Intron number	5′ splice site	Putative branch site	3′ splice site	Size (nt)
Human	SCN4A	21	AG atgagtat...tcaacctgac	-ccactatac	TT	822
Human	SCN5A	25	AG atacgtgg...ggcctctgag	-tctttgcac	TT	680
P. chrysogenum	xylP	7	GG ataagtac...actagctaac	agcctctcac	AG	50
A. nidulans	xlnC	8	GG ataagttt...tcacactgac	aaccccaaac	AG	52
A. tubingensis	xlnC	7	GG ataagtac...gatagctgac	agcccctcac	AG	52
AT-AC II consensus			RG ataagt	ctrac	yac	
GT-AG consensus			AG gtaagt	ynytray	y_{11}nyag	GT
AT-AC I consensus			atatccty	tccttrac	yac	

[a]SCN4A: voltage-gated skeletal muscle (type IV) sodium channel α subunit (George et al., 1993; size determined from our unpublished sequencing data); SCN5A: voltage-gated cardiac sodium channel α subunit (Wang et al., 1996; size determined from our unpublished sequencing data); xylP: xylanase gene from *P. chrysogenum* (Haas et al., 1993); xlnC: genes encoding the 34-kDa xylanase in *A. nidulans* (MacCabe et al., 1996) or *A. tubingensis* (D. Ramón, pers. comm.). Intron sequences are shown in lower case, flanking exon sequences in upper case, invariant intron sequence elements in white letters on black background, and less conserved sequences in black letters on grey background. Dots indicate sequences that are either not shown or not known; the dash in the first two sequences denotes a stretch of sequence that is not shown. The consensus sequence is shown at the bottom, with the GT-AG and AT-AC I consensus sequences shown for comparison. Underlined adenosines indicate the branch points; r = purine; y = pyrimidine.

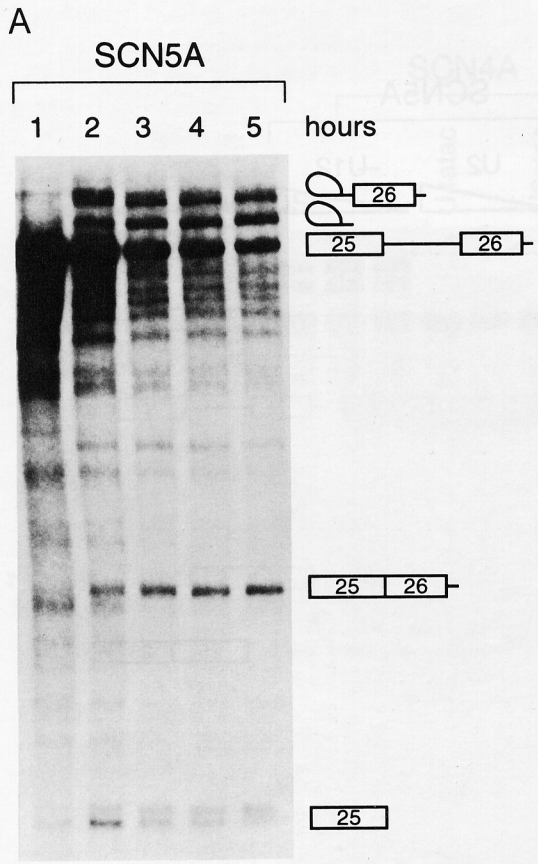
genita (Plassart et al., 1994; Hoffman, 1995). One of the introns present in each of three homologous xylanase genes from filamentous fungi also appears to be an AT-AC II intron (Table 2) (Haas et al., 1993; MacCabe et al., 1996; D. Ramón, pers. comm.).

In the present study, we cloned portions of the human SCN4A and SCN5A genes spanning the AT-AC II introns 21 and 25, respectively, and their flanking exons, and established conditions for splicing the corresponding transcripts in vitro. Using oligonucleotide-directed RNase H cleavage or annealing of complementary 2′-O-methyl oligonucleotides to inhibit the activity of major and minor U snRNPs, we have compared the snRNA requirements for splicing of AT-AC II introns to those for splicing of AT-AC I and conventional GT-AG introns.

RESULTS

Splicing of SCN5A AT-AC II intron 25 pre-mRNA in HeLa cell nuclear extract

An AT-AC II intron splicing substrate derived from the SCN5A gene was constructed from intron 25 and both flanking exons, followed by 6 nt of the downstream intron (Fig. 1). The downstream conventional 5′ splice site was included because of potential exon-definition interactions that might improve splicing efficiency (reviewed in Berget, 1995). The SCN5A pre-mRNA was spliced with moderate efficiency in the nuclear extract (Fig. 1A). A lariat intron 25–exon 26 intermediate and a released intron lariat were generated and migrated above the pre-mRNA, due to their



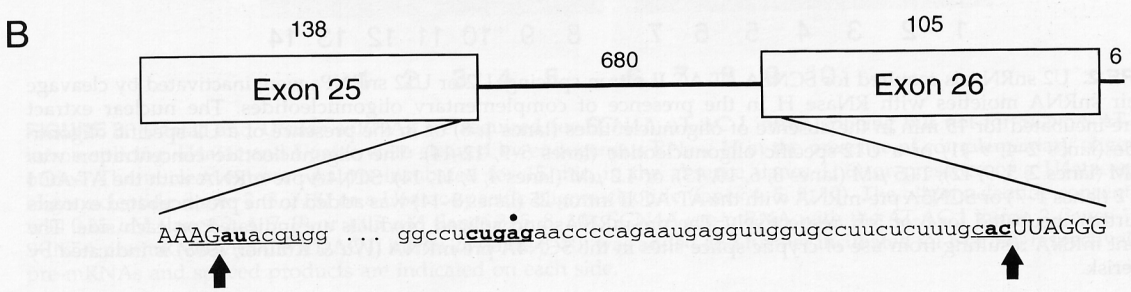


FIGURE 1. In vitro splicing of SCN5A AT-AC II intron 25. **A:** Time course of in vitro splicing of the SCN5A pre-mRNA in HeLa cell nuclear extract. Capped, in vitro-transcribed, ^{32}P-labeled SCN5A pre-mRNA was incubated under optimized splicing conditions for the indicated times and analyzed by urea-PAGE and autoradiography. Splicing efficiency did not improve beyond 5 h (data not shown). The structures and electrophoretic mobilities of the substrate, intermediates, and products are indicated on the right. **B:** Schematic structure of the SCN5A pre-mRNA substrate, showing the nucleotide sequence surrounding the exon–intron boundaries. The exon and intron sizes in nucleotides are indicated at the top. Exon sequences are in upper case, intron sequences are in lower case, conserved sequence elements are underlined, and the putative branchpoint adenosine is indicated by a dot. The actual 5′ and 3′ splice sites, determined by sequencing of RT-PCR products of in vitro splicing, are indicated by solid arrows.

nonlinear nature. The nucleotide sequence of the gel-purified spliced mRNA was determined by reverse transcription-PCR (RT-PCR), and showed unambiguously that splicing was accurate and that cleavage occurred immediately 5′ to the AU and 3′ to the AC dinucleotides at the splice sites (Fig. 1B).

Essentially identical results were obtained with a pre-mRNA comprising the closely related intron 21 and flanking exons from the SCN4A gene (data not shown). For both of these introns, a GU dinucleotide is present four bases downstream of the AU 5′ splice site, and an AG dinucleotide is present four bases downstream of the AC 3′ splice site. However, none of several RT-PCR products that were sequenced represented splicing via these GU-AG sites. In vitro splicing via the AU-AC splice sites is fully consistent with the sequences of cloned SCN5A and SCN4A cDNAs (George et al., 1993; Wang et al., 1996).

Role of U2 and U12 snRNAs in splicing of AT-AC II versus AT-AC I introns

Oligonucleotide-directed RNase H cleavage has been used to test the involvement of U1, U2, U4, and U6 snRNAs in the GT-AG splicing pathway (Krämer et al., 1984; Black et al., 1985; Krainer & Maniatis, 1985; Berget & Robberson, 1986; Black & Steitz, 1986), as well as the requirement for U12 snRNA (Wu & Krainer, 1996) and U4atac snRNA (Tarn & Steitz, 1996b) in the AT-AC I intron splicing pathway. In this study, we first used oligonucleotide-directed RNase H cleavage to test the requirement for U2 and U12 snRNAs in AT-AC II intron splicing. When these snRNAs were cleaved by RNase H and complementary oligonucleotides, striking differences were observed in the behavior of the SCN5A substrate with an AT-AC II intron and the SCN4A control substrate with an AT-AC I intron (Fig. 2). With the

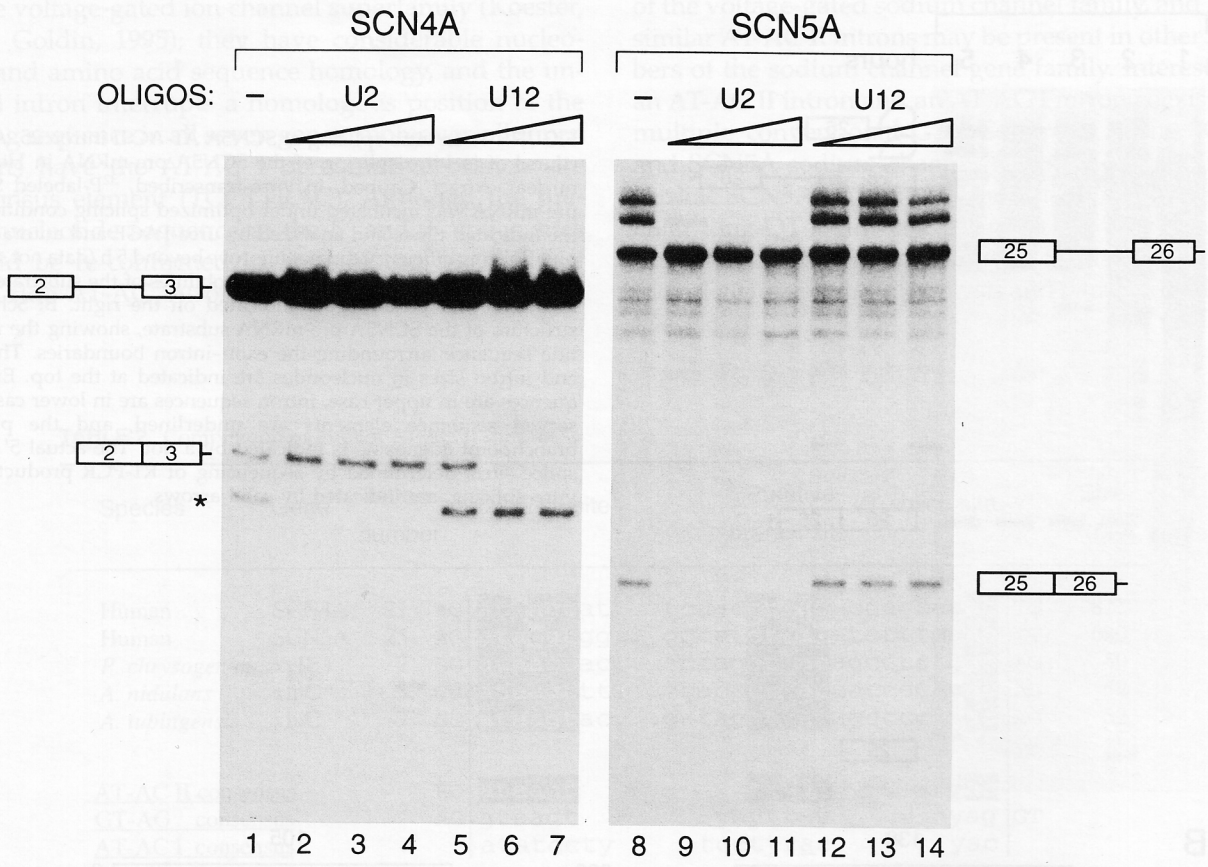


FIGURE 2. U2 snRNA is required for SCN5A AT-AC II intron splicing. U2 or U12 snRNPs were inactivated by cleavage of their snRNA moieties with RNase H in the presence of complementary oligonucleotides. The nuclear extract was pre-incubated for 15 min in the absence of oligonucleotides (lanes 1, 8) or in the presence of a U2-specific oligonucleotide (lanes 2–4, 9–11) or a U12-specific oligonucleotide (lanes 5–7, 12–14). The oligonucleotide concentration was 0.25 μM (lanes 2, 5, 9, 12), 1.05 μM (lanes 3, 6, 10, 13), or 4.2 μM (lanes 4, 7, 11, 14). SCN4A pre-mRNA with the AT-AC I intron 2 (lanes 1–7) or SCN5A pre-mRNA with the AT-AC II intron 25 (lanes 8–14) was added to the pre-incubated extracts and further incubated for 6 or 5 h, respectively. The pre-mRNAs and spliced products are indicated on each side. The aberrant mRNA resulting from use of cryptic splice sites in the SCN4A pre-mRNA (Wu & Krainer, 1996) is indicated by an asterisk.

SCN5A pre-mRNA (lanes 8–14), the U2-specific oligonucleotide inhibited splicing completely (compare lanes 9–11 with 8), whereas the U12-specific oligonucleotide had no effect (lanes 12–14). In contrast, with the control SCN4A pre-mRNA (lanes 1–7), as described previously (Wu & Krainer, 1996), the U12 oligonucleotide inhibited splicing completely (lanes 5–7), whereas the U2 oligonucleotide had no effect (lanes 2–4). U2 and U12 oligonucleotides also had opposite effects on splicing via GT-AG cryptic splice sites in the SCN4A pre-mRNA, which were the reciprocal of the effects on splicing via the AT-AC I splice sites, as expected (Wu & Krainer, 1996). We conclude that U2 snRNA, rather than U12 snRNA, is required for splicing of the SCN5A AT-AC II intron 25.

U4atac and U6atac snRNAs are required for SCN4A AT-AC I intron 2 splicing but not for SCN5A AT-AC II intron 25 splicing

The U4atac and U6atac minor snRNAs were identified recently, sequenced, and shown to be required for splicing of the P120 AT-AC I intron in vitro (Tarn & Steitz, 1996b). We tested the requirement for these snRNAs in SCN4A AT-AC I intron 2 splicing by RNase H cleavage, and confirmed that disruption of these snRNAs inhibits processing of an AT-AC I intron, but not of the intron defined by cryptic GT-AG splice sites (Fig. 3, lanes 2–5). In contrast, disruption of U4atac and U6atac snRNAs did not inhibit splicing of the SCN5A AT-AC II intron 25 splicing (lanes 7–10). We

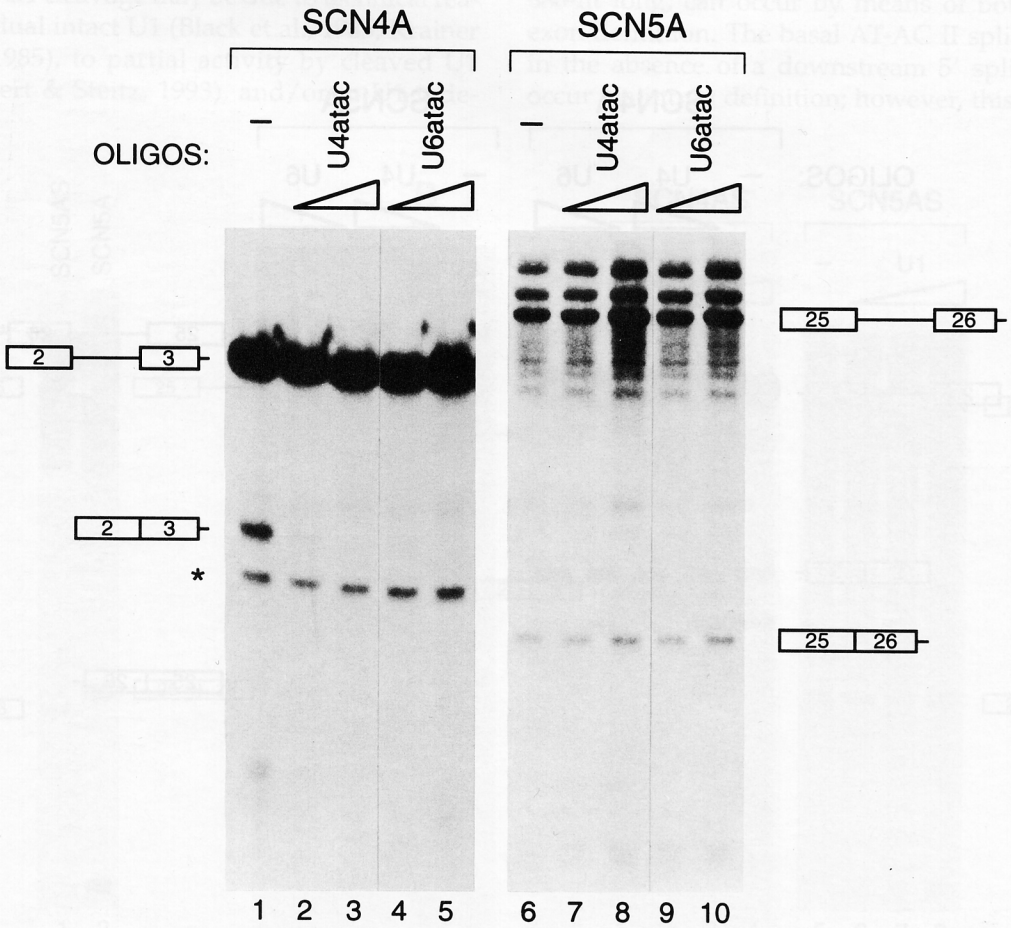


FIGURE 3. U4atac and U6atac snRNAs are required for SCN4A AT-AC I intron splicing but not for SCN5A AT-AC II intron splicing. U4atac and U6atac were cleaved by endogenous RNase H in the presence of complementary oligonucleotides. The nuclear extract was pre-incubated for 15 min in the absence (lanes 1, 6) or presence of a U4atac-specific oligonucleotide (lanes 2, 3, 7, 8) or a U6atac-specific oligonucleotide (lanes 4, 5, 9, 10). The oligonucleotide concentration was 0.25 μM (lanes 2, 4, 7, 9) or 1.05 μM (lanes 3, 5, 8, 10). SCN4A pre-mRNA with the AT-AC I intron 2 (lanes 1–5) or SCN5A pre-mRNA with the AT-AC II intron 25 (lanes 6–10) was added and further incubated for 6 or 5 h, respectively. The pre-mRNAs and spliced products are indicated on each side.

conclude that U4atac and U6atac snRNAs are required for splicing of SCN4A AT-AC I intron 2 splicing, but not for SCN5A AT-AC II intron 25 splicing.

U4 and U6 snRNAs are required for SCN5A AT-AC II intron 25 splicing

Because U4atac and U6atac snRNAs are not required for SCN5A intron 25 splicing, we tested whether this reaction depends on U4 and U6 snRNAs. U4atac and U6atac snRNAs are thought to play analogous roles in the AT-AC I spliceosome to those of U4 and U6 snRNAs in the GT-AG spliceosome (Nilsen, 1996; Tarn & Steitz, 1996b). Oligonucleotide-directed RNase H cleavage of U4 and U6 snRNAs resulted in inhibition of SCN5A intron 25 splicing (Fig. 4, lanes 6–10). As expected, cleavage of these snRNAs had no effect on the control SCN4A intron 2 reaction, but did inhibit conventional splicing via the cryptic GT-AG splice sites (lanes 1–5). We conclude that U4 and U6 snRNAs are required for splicing of the SCN5A AT-AC II intron 25.

U1 snRNA requirement for AT-AC II intron splicing

The U1 snRNA requirement for the basal AT-AC II intron splicing reaction cannot be tested directly with the above SCN5A intron 25 substrate, because of the presence of the downstream conventional 5′ splice site, which may bind U1 snRNA, resulting in exon-definition interactions that stimulate splicing of the preceding intron (Robberson et al., 1990; Kuo et al., 1991; Wu & Krainer, 1996). Therefore, we constructed a shorter substrate, designated SCN5AS, lacking the downstream 5′ splice site. The SCN5AS pre-mRNA was spliced accurately in the nuclear extract, albeit with reduced

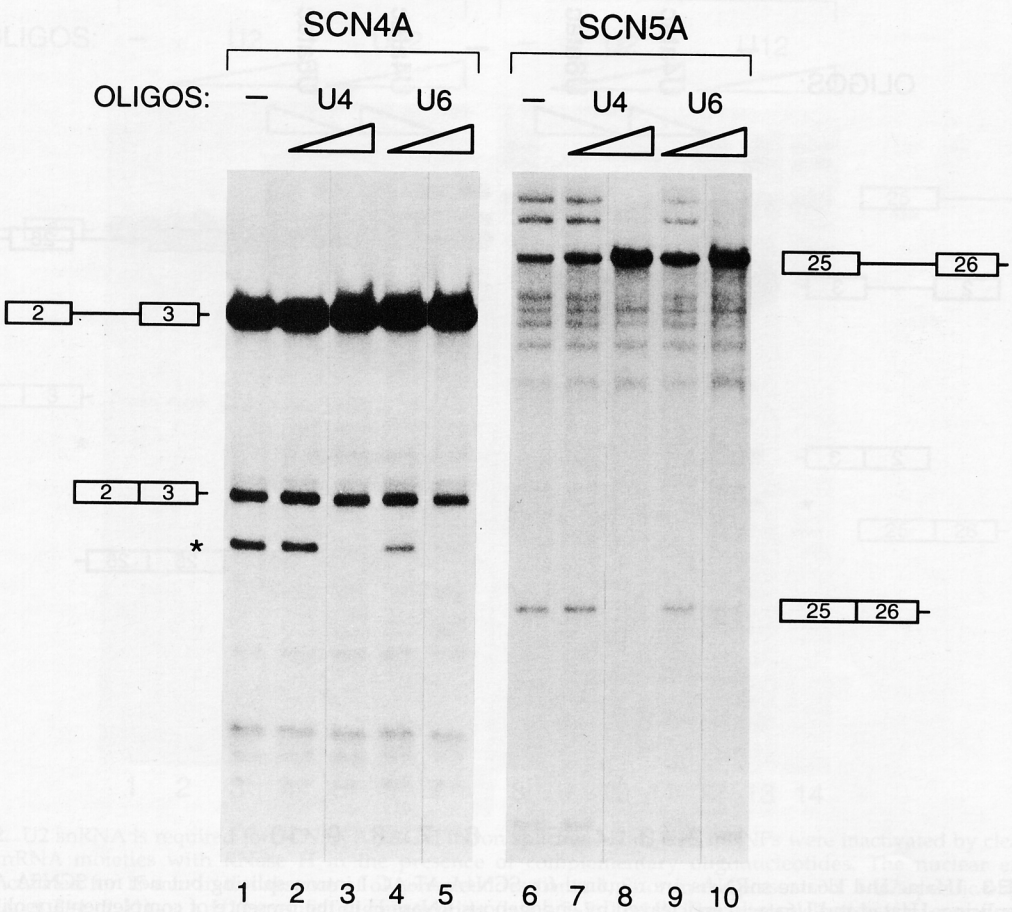


FIGURE 4. U4 and U6 snRNAs are required for SCN5A AT-AC II intron splicing. U4 and U6 snRNAs were inactivated by RNase H and complementary oligonucleotides. The nuclear extract was pre-incubated for 15 min in the absence (lanes 1, 6) or presence of U4-specific (lanes 2, 3, 7, 8) or U6-specific (lanes 4, 5, 9, 10) oligonucleotides. The oligonucleotide concentration was 0.25 μM (lanes 2, 4, 7, 9) or 4.2 μM (lanes 3, 5, 8, 10). SCN4A pre-mRNA with the AT-AC I intron 2 (lanes 1–5) or SCN5A pre-mRNA with the AT-AC II intron 25 (lanes 6–10) was added and further incubated for 6 or 5 h, respectively. The pre-mRNAs and spliced products are indicated on each side.

efficiency compared to the SCN5A pre-mRNA (Fig. 5A). The reduced efficiency likely reflects the loss of stimulatory exon-definition interactions between components bound at the 3′ splice site of the AT-AC II intron 25 and U1 snRNP bound at the 5′ splice site of the GT-AG intron 26. When U1 snRNA was cleaved by RNase H in the presence of a complementary oligonucleotide, SCN5AS splicing was inhibited partially, apparently in the second step (Fig. 5B, compare lanes 6–8 to lane 5). As a control, we analyzed an SCN4A pre-mRNA (SCN4AS) comprising the AT-AC I intron 2, but lacking the downstream 5′ splice site of the GT-AG intron 3. Splicing of the SCN4AS pre-mRNA was enhanced upon cleavage of U1 snRNA, whereas splicing via the cryptic GT-AG splice sites was inhibited (compare lanes 2–4 to lane 1). Basal AT-AC I splicing is U1 independent (Wu & Krainer, 1996), and the observed stimulation upon U1 cleavage (Tarn & Steitz, 1996a) probably reflects inhibition of the competing cryptic GT-AG pathway. Incomplete inhibition of the AT-AC II reaction upon U1 cleavage may be due to technical reasons, i.e., residual intact U1 (Black et al., 1985; Krainer & Maniatis, 1985), to partial activity by cleaved U1 snRNA (Seiwert & Steitz, 1993), and/or to U1-independent splicing (Crispino et al., 1994; Tarn & Steitz, 1994). With the SCN5A pre-mRNA, U1 cleavage resulted in more severe inhibition of splicing than with the SCN5AS pre-mRNA (data not shown), suggesting that at least two molecules of U1 snRNP are involved in this reaction, one for binding to the 5′ splice site of the AT-AC II intron 25, and one to the 5′ splice site of the GT-AG intron 26.

The exon definition model proposed that, in pre-mRNAs with large introns, i.e., most vertebrate introns, spliceosome assembly initially involves recognition of pairs of splice sites flanking each small exon; subsequently, the neighboring exons are juxtaposed for splicing catalysis (Berget, 1995). When introns are small, such as in yeast (Guthrie, 1991) and *Drosophila* (Mount et al., 1992), spliceosome assembly proceeds by intron definition, i.e., the 5′ and 3′ splice sites from the same intron are recognized and brought together (Talerico & Berget, 1994). The above experiments show that splicing of the SCN5A AT-AC II intron, which is 680-nt long, can occur by means of both intron and exon definition. The basal AT-AC II splicing reaction, in the absence of a downstream 5′ splice site, must occur via intron definition; however, this reaction was

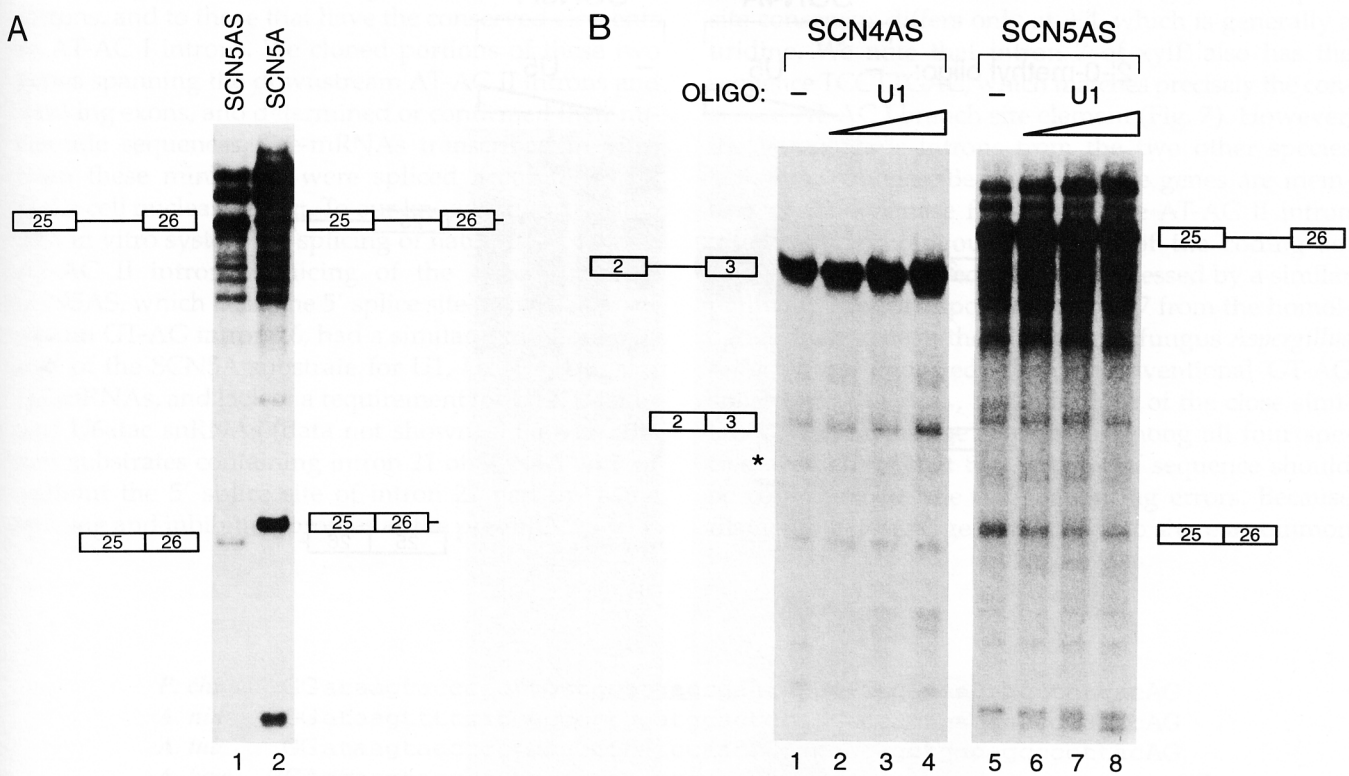


FIGURE 5. **A:** Exon definition interaction between an AT-AC II intron and the downstream GT-AG intron. Equal amounts of SCN5AS (lane 1) or SCN5A (lane 2) pre-mRNAs with the AT-AC II intron 25 were spliced in vitro for 5 h. The pre-mRNA substrates and the spliced mRNA products are indicated schematically on each side. The SCN5AS pre-mRNA lacks the downstream intron 26 5′ splice site (6 nt) present in the SCN5A pre-mRNA. **B:** U1 snRNA is involved in SCN5AS AT-AC II intron splicing. Nuclear extract was pre-incubated for 15 min in the absence (lanes 1, 5) or presence (lanes 2–4, 6–8) of a U1-specific oligonucleotide. The oligonucleotide concentration was 0.25 μM (lanes 2, 6), 1.05 μM (lanes 3, 7) or 4.2 μM (lanes 4, 8). SCN4AS pre-mRNA with the AT-AC I intron 2 (lanes 1–4) or SCN5AS pre-mRNA with the AT-AC II intron 25 (lanes 5–8) was added and further incubated for 6 or 5 h, respectively.

stimulated strongly by intact U1 snRNA and a downstream GT-AG 5′ splice site, presumably through exon definition bridging interactions involving U1 and U2 snRNPs, SR proteins, and U2AF (Kuo et al., 1991; Hoffman & Grabowski, 1992; Wu & Maniatis, 1993).

U5 is required for splicing of SCN4A AT-AC I intron 2 and SCN5A AT-AC II intron 25

The role of U11 and U5 snRNAs in the AT-AC I and AT-AC II splicing pathways could not be examined by the oligonucleotide-directed RNase H cleavage approach, because these snRNAs are resistant to cleavage within the native snRNP particles (Montzka & Steitz, 1988; Black & Pinto, 1989). Instead, we used the alternative approach of inhibiting the activity of these snRNPs by annealing 2′-O-methyl oligonucleotides, which bind tightly to complementary RNAs and do not confer sensitivity to RNase H (Inoue et al., 1987; Blencowe et al., 1989). This approach has been used recently to demonstrate U12, U5, and U6atac snRNA involvement in AT-AC I intron splicing (Tarn & Steitz, 1996a, 1996b). Incubation with a U5-specific 2′-O-methyl oligonucleotide inhibited splicing of both SCN4A and SCN5A pre-mRNAs (Fig. 6). As expected, it also inhibited splicing via the cryptic GT-AG splice sites in the SCN4A pre-mRNA. With the SCN5A pre-mRNA, the second step of splicing appeared to be inhibited preferentially, resulting in accumulation of the intermediates (lane 7). U5 snRNP is known to be involved in both steps of the conventional GT-AG intron splicing pathway (Newman & Norman, 1992), and in at least the first step of the AT-AC I intron splicing pathway (Tarn & Steitz, 1996a). The inhibition we observed by debilitation of U5 snRNP appears to be specific, because a control 2′-O-methyl oligonucleotide, assayed under identical conditions, had no effect on splicing of either GT-AG, AT-AC I, or AT-AC II introns (data not shown). We conducted similar experiments with a U11-specific 2′-O-methyl oligonucleotide and observed inhibition of both SCN4A and SCN5A pre-mRNA splicing (data not shown). However, although the SCN4A inhibition data are consistent with the proposal that U11 snRNA is involved in AT-AC I intron

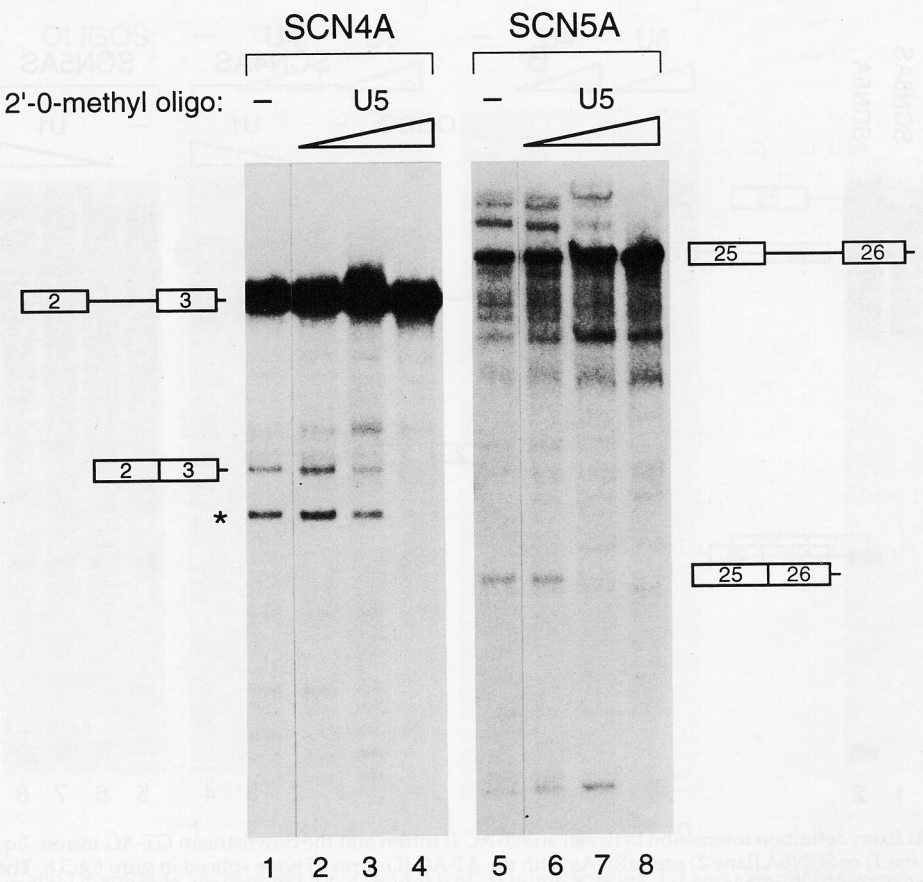


FIGURE 6. AT-AC I intron splicing and AT-AC II intron splicing require U5 snRNA. Nuclear extract was pre-treated in the absence (lanes 1, 5) or presence (lanes 2–4, 6–8) of U5-specific 2′-O-methyl oligonucleotide at a concentration of 0.25 μM (lanes 2, 6), 1.05 μM (lanes 3, 7), or 4.2 μM (lanes 4, 8) under splicing conditions for 10 min. SCN4A pre-mRNA with the AT-AC I intron 2 (lanes 1–4) or SCN5A pre-mRNA with the AT-AC II intron 25 (lanes 5–8) was added and further incubated for 6 or 5 h, respectively. Pre-mRNAs and spliced mRNAs are indicated schematically on each side.

splicing (Hall & Padgett, 1994; Tarn & Steitz, 1996a), we observed comparable inhibition of β-globin GT-AG intron 1 splicing in the same range of oligonucleotide concentration (data not shown). U11 snRNA is not known to be involved in the conventional GT-AG splicing pathway, so perhaps the oligonucleotides fortuitously bound to, and inactivated, another component required for splicing. Additional experiments will be required to understand the role of U11 in these splicing pathways.

DISCUSSION

Two functionally distinct subclasses of noncanonical AT-AC introns

Not all introns that begin with AT and end with AC have the same sequence features at the splice sites and presumptive branch site (Wu & Krainer, 1996). Intron 21 of SCN4A and intron 25 of SCN5A, which interrupt homologous positions of these sodium channel genes, are distinct from other AT-AC introns in that the sequences at their 5′ splice sites and branch sites diverge considerably from those of other AT-AC introns. We refer to these divergent introns as AT-AC II introns, and to those that have the conserved elements as AT-AC I introns. We cloned portions of these two genes spanning the downstream AT-AC II introns and flanking exons, and determined or confirmed their nucleotide sequences. Pre-mRNAs transcribed in vitro from these minigenes were spliced accurately in a HeLa cell nuclear extract. To our knowledge, this is the first in vitro system for splicing of naturally occurring AT-AC II introns. Splicing of the short substrate SCN5AS, which lacks the 5′ splice site from the downstream GT-AG intron 26, had a similar requirement as that of the SCN5A substrate for U1, U2, U4, U6, and U5 snRNAs, and lack of a requirement for U12, U4atac, and U6atac snRNAs (data not shown). Likewise, the two substrates containing intron 21 of SCN4A with or without the 5′ splice site of intron 22 had the same splicing and inhibition profiles as the pre-mRNAs with intron 25 of SCN5A (data not shown). Thus, these two closely related AT-AC II introns are functionally distinct from AT-AC I introns.

The two human AT-AC II introns known to date interrupt the same position of the coding sequence of two paralogous members of the sodium channel gene family, SCN4A and SCN5A, which remarkably also have an AT-AC I intron and multiple conventional GT-AG introns. Similar AT-AC II introns are expected to exist in other members of this gene family, whose genomic structures are still unknown. Three members of a subfamily of xylanase genes from filamentous fungi also have what appears to be an AT-AC II intron (Table 2; Fig. 7). Intron 7 of xylP in *Penicillium chrysogenum* (Haas et al., 1993), intron 8 of xlnC in *Aspergillus nidulans* (MacCabe et al., 1996), and intron 7 of xlnC in *Aspergillus tubingensis* (D. Ramón, pers. comm.) have 5′-AT and 3′-AC splice sites. However, they lack the conserved ATCCT sequence from position +3 to +7 of the 5′ splice site. All three introns have putative branch site sequences (Table 2) with complementarity to the U2 snRNA sequences that function in recognition of GT-AG intron branch sites (Parker et al., 1987). The 5′ splice sites of these fungal introns match precisely the metazoan GT-AG consensus 5′ splice site, AGgtaagt, from +2 to +6 and at −1; the budding yeast 5′ splice site consensus differs only at +4, which is generally a uridine. We note that intron 7 of xylP also has the sequence TCCTTGAC, which matches precisely the conserved AT-AC I branch site element (Fig. 7). However, the homologous introns from the two other species lack this sequence. Because all three genes are members of the xylanase family and the AT-AC II intron interrupts homologous positions of the coding sequence, they are expected to be processed by a similar pathway. The corresponding intron 7 from the homologous *xynA* gene in the filamentous fungus *Aspergillus kawachii* was reported to have conventional GT-AG splice sites (Ito et al., 1992); in light of the close similarity of the xylanase sequences among all four species, we believe that the *A. kawachii* sequence should be confirmed to rule out sequencing errors. Because filamentous fungal genera appear to share a common

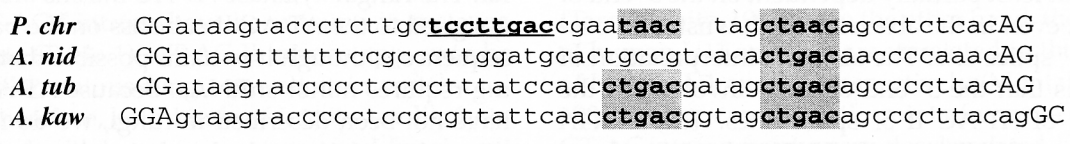


FIGURE 7. Sequence alignment of homologous xylanase introns from four filamentous fungi. *P. chrysogenum* xylP gene intron 7 (P. chr) (Haas et al., 1993); *A. nidulans* xlnC gene intron 8 (A. nid) (MacCabe et al., 1996); *A. tubingensis* xlnC gene intron 7 (A. tub) (J. Catalán, J.A. Pérez-Gonzalez, & D. Ramón, unpubl. results); *A. kawachii* xynA gene intron 7 (A. kaw) (Ito et al., 1992). Exon sequences are shown in upper case and intron sequences in lower case. Putative branch sites that fit the CTRAC consensus are indicated by shading. The bold and underlined octamer in the top sequence indicates a perfect match to the AT-AC I branch site consensus (Table 1). At least the top three introns appear to be of the AT-AC II subclass.

splicing mechanism (Gurr et al., 1987), it will be interesting to see whether *Schizosaccharomyces pombe* also has AT-AC II introns. Whether AT-AC I or AT-AC II introns exist in *Saccharomyces cerevisiae* will be known soon, as annotation of its complete genome sequence (Goffeau et al., 1996) is conducted.

Comparison between AT-AC and GT-AG spliceosomes

Splicing of GT-AG and AT-AC I introns involves similar two-step trans-esterification reactions, assembly of spliceosomal complexes, and the generation of lariat intermediates (Tarn & Steitz, 1996a). The recognition of the critical sites on the respective pre-mRNAs involves distinct sets of snRNAs, but appears to involve analogous base pairing interactions (reviewed in Nilsen, 1996). Thus, U12 snRNA base pairs to the AT-AC I intron branch site element, whereas U2 snRNA performs this role in GT-AG spliceosomes (Hall & Padgett, 1996; Tarn & Steitz, 1996a); U11 snRNA is present in the AT-AC I spliceosome (Tarn & Steitz, 1996a) and has been proposed to base pair to the 5′ splice site, a role performed by U1 snRNA in the GT-AG spliceosome (Hall & Padgett, 1994). U12 and a fraction of U11 form a di-snRNP particle (Wassarman & Steitz, 1992), and likewise, interactions between a small fraction of U2 and U1 snRNPs have been detected in *Xenopus* (Mattaj et al., 1986). Two novel snRNAs, U4atac and U6atac, are part of a di-snRNP particle and show some sequence similarity to U4 and U6 snRNAs, which are also present in a di-snRNP particle (Tarn & Steitz, 1996b). The U5 snRNA appears to be shared between AT-AC I and GT-AG pathways (Fig. 6; Tarn & Steitz, 1996a); however, at least seven closely related U5 snRNA variants have been identified (Sontheimer & Steitz, 1992), and it is not known whether specific variants function in each pathway. A triple snRNP particle involving U4atac, U6atac, and U5 snRNAs—by analogy to the U4/U6·U5 tri-snRNP—has not been described yet. It remains possible that there are introns that are spliced by hybrid pathways, involving combinations of snRNA and/or protein factors from the GT-AG and AT-AC I pathways.

Splicing of the two AT-AC II introns we studied appears to be at least partially dependent on the 5′ end of U1 snRNA, even in the absence of a downstream conventional 5′ splice site. The simplest interpretation of this result is that the 5′ end of U1 is involved in the recognition of AT-AC II 5′ splice sites. The SCN4A intron 21 and SCN5A intron 25 5′ splice sites have good complementarity to the 5′ end of U1 snRNA, except that position 1, which usually base pairs with C_8 of U1, is an A instead of a G. On the other hand, AT-AC II introns may be more strongly dependent on U6 snRNA than on U1 snRNA for recognition and/or proofreading of the 5′ splice site (Kandels-Lewis & Séraphin, 1993; Lesser & Guthrie, 1993; Sontheimer & Steitz, 1993; Tarn & Steitz, 1994; Crispino & Sharp, 1995). Positions 2, 5, and 6 of the 5′ splice sites of SCN4A intron 21 and SCN5A intron 25 have good complementarity to A_{45}, C_{42}, and A_{41}, respectively, within the highly conserved ACAGAG box of U6 snRNA (Sun & Manley, 1995; Hwang & Cohen, 1996). Another possibility is that U11 snRNA may be involved in recognition of AT-AC II 5′ splice sites; residues 8–13 of U11 are complementary at five of six or four of six positions to the 5′ splice sites of SCN4A intron 21 and SCN5A intron 25, respectively. Our 2′-O-methyl oligonucleotide inhibition data are consistent with, but do not prove, this possibility. These scenarios are not mutually exclusive: each of these three snRNAs could conceivably recognize the AT-AC II 5′ splice site in a sequential fashion, as a proofreading mechanism.

Intron 21 of SCN4A and intron 25 of SCN5A have presumptive branch site elements that conform to the weak consensus for GT-AG intron branches, YNYTRAY, at six of seven positions. The distance between the presumptive branch adenosine and the 3′ splice site of these introns is 21 and 37 nt, respectively. This distance is within the norm for GT-AG introns, whereas AT-AC I branch sites, which are highly conserved, are usually closer to the 3′ splice site (9–13 nt, Table 1). The SCN4A and SCN5A AT-AC II introns have runs of pyrimidines upstream of their 3′ splice sites, a feature of most GT-AG introns, but not of AT-AC I introns. The highly conserved AT-AC I branch site, TCCTTRAY, matches the loose metazoan GT-AG consensus branch.

Our results show that processing of AT-AC II introns and GT-AG introns requires the same set of snRNAs. This observation is consistent with previous in vivo second-site suppression experiments, which showed that, in at least some conventional GT-AG introns, G to A mutation at the first position and G to C mutation at the last position suppress each other, both in yeast (Parker & Siliciano, 1993; Chanfreau et al., 1994) and human (Deirdre et al., 1995). However, because spliceosomes contain numerous proteins that are still not completed defined, we cannot conclude at the present time that the splicing mechanism of GT-AG and AT-AC II introns is identical in all respects. Moreover, very few natural AT-AC II introns are known so far. The fungal xylanase AT-AC introns are tentatively assigned to the AT-AC II subclass on the basis of their sequence, although it remains possible that their splicing requires minor snRNAs. Because AT-AC I introns have not been described in fungi, we do not know if they exist and, if so, whether their splice site and branch sequence elements are identical to those of metazoans. Although additional AT-AC II introns will undoubtedly be identified, this type of intron appears to belong to a rare subclass, even less common than AT-AC I introns.

Is there a direct non-Watson–Crick interaction between the intron ends?

Almost all naturally occurring introns have 5′ splice site GT and 3′ splice site AG boundaries. The importance of the intron ends is underscored by the numerous genetic diseases caused by splice site mutations (Krawczak et al., 1992; Nakai & Sakamoto, 1994). Through mutational analysis of intron ends, it has been observed that G → A and G → C mutations at the first and last positions, respectively, of several GT-AG introns suppress each other to restore splicing. This type of result has been interpreted to suggest non-Watson–Crick base pairing between the first and last bases of the intron (Parker & Siliciano, 1993; Chanfreau et al., 1994; Deirdre et al., 1995). Because the individual mutations allow the first step of splicing, this interaction is presumed to occur after formation of the lariat–exon intermediate. The first nucleotide of the intron is covalently joined to the branchpoint A via a 2′-5′ phosphodiester linkage, but this would still allow for hydrogen bonding between the first and last bases, by looping out the usually short segment of RNA between the branchpoint and the 3′ splice site. This presumptive interaction might serve as a proofreading step. Alternative interpretations of the suppression data are that the first and last nucleotides are recognized by separate but synergistic factors, or that certain combinations of nucleotides at these positions are sterically excluded through proofreading steps (Luukkonen & Séraphin, 1997). The discovery of natural AT-AC introns suggested yet another explanation for the original suppression experiments: mutation of the first and last bases could theoretically have resulted in a shift from the conventional to the AT-AC splicing pathway. This possibility was perhaps unlikely, because of the conservation of the unique 5′ splice site and branch site elements of AT-AC I introns, but it could not be discounted a priori. Our observations with naturally occurring AT-AC II introns, which resemble GT-AG introns except for the first and last bases, show that these substrates utilize the major spliceosomal snRNAs, rather than the minor AT-AC snRNAs. Therefore, the same is likely to be the case for the double-mutant GT-AG introns.

Evolutionary origin of AT-AC introns

To date, AT-AC I introns have been found in metazoa and in plants (Table 1). AT-AC II introns appear to be present in filamentous fungi, as well as in vertebrates (Table 2). Therefore, both kinds of introns are likely to be distributed widely among metazoa and fungi, and perhaps through the entire eukaryotic lineage. The two examples of AT-AC intron sequences found in plant genomes are likely of the AT-AC I type, based on the conservation of the 5′ splice site and presumptive branch site sequences. If this is indeed the case, it appears that AT-AC I introns arose more than a billion years ago (Wu et al., 1996). At the present time, it is not certain that the two described *Arabidopsis* introns are bona fide AT-AC I introns: the AtG5 intron ends with AA, rather than AC, and it has a divergent presumptive branch site sequence; the ArLIM15 intron, as defined by Wu et al. (1996), strikingly has all the sequence features of an AT-AC I intron, but the splice junctions indicated in the original sequence (Sato et al., 1995) are different. The assignment of splice sites by Wu et al. (1996) is supported by alignment of the proposed coding sequences with other recA-like proteins (Sato et al., 1995). At the very least, because AT-AC I introns are found at the homologous position in genes of the voltage-gated sodium and calcium channel α-subunit superfamily, their existence predates the divergence of the sodium and calcium channel genes about 700 million years ago (Hille, 1992). There is no intron at the corresponding position of the *Drosophila* sodium channel α-subunit gene *para* (George et al., 1993). Voltage-gated sodium and calcium channels have four internal homologous domains that are thought to arise from two rounds of duplication from a single ancestral gene (Koester, 1991; Hille, 1992; Goldin, 1995). Although voltage-gated sodium channel genes have an AT-AC II intron between internal domains 3 and 4, the calcium channel genes do not have any introns at the homologous position. The AT-AC I intron of the sodium and calcium channel genes is found within domain 1. Sequence alignment of these domains shows that the two AT-AC introns are not located at the same position within each duplicated domain; moreover, no introns are present at the corresponding positions of the other two duplicated domains (data not shown). These observations suggest that the AT-AC I and AT-AC II introns in the ion channel genes arose independently.

The recent discovery of AT-AC pre-mRNA introns raises the question of their evolutionary relationship with conventional GT-AG pre-mRNA introns. There has been speculation that AT-AC I introns and the minor snRNAs involved in their processing represent a molecular fossil (Mount, 1996). AT-AC II introns may represent an intermediate type of intron between AT-AC I and GT-AG introns. The sequence similarity between human U4atac and U4, and between U6atac and U6 is about 40%, whereas yeast and human U6 are more similar to each other than human U6 is to human U6atac (Tarn & Steitz, 1996b). On the other hand, the sequences of human U11 and U12 are unrelated to the sequences of human U1 and U2, respectively (Montzka & Steitz, 1988). From a mechanistic point of view, AT-AC II and GT-AG introns may be very closely related, because they employ a common set of snRNAs, and point mutation of the first and last nucleotides of a GT-AG intron can generate a functional AT-AC II intron experimentally. Whether AT-AC I and GT-AG introns are derived from a common ancestor is not known. Although the respec-

tive splicing pathways are closely related, generating similar lariat molecules and involving snRNAs that have apparently similar roles, each pathway makes use of several unique snRNAs.

It is puzzling that AT-AC II introns are not more prevalent. Although two point mutations are required to convert a GT-AG intron to an AT-AC II intron with the same exon–intron boundaries, any single mutation that inactivates a conventional 5′ or 3′ splice site would be expected to activate pairs of cryptic AT-AC II splice sites occasionally. However, to our knowledge, of the many reported mutations that result in activation of new splice site pairs, none involve AT-AC boundaries. It could be argued that, because GT-AG introns appear to splice more efficiently than introns with natural or experimentally generated AT-AC boundaries, the higher efficiency could confer a selective advantage to GT-AG introns. However, efficient splicing is not always required for gene expression and, in fact, inefficient splicing is often advantageous. Thus, many pre-mRNAs with GT-AG introns splice inefficiently, which can be crucial for the proper expression of certain viral and cellular genes.

An interesting question is how an AT-AC I intron commits to splicing via the AT-AC I pathway, rather than the GT-AG pathway. The previous analysis of GT-AG to AT-AC mutant introns, and the present description of natural AT-AC II introns, indicate that the boundary nucleotides are not responsible for the lack of commitment to the standard GT-AG pathway. The AT-AC I branch site element, which is highly conserved and specifies base pairing to U12 snRNA, may be important. However, its sequence fits the standard GT-AG branch site consensus as well, and potentially it can base pair to the correct region of U2. Perhaps the closer proximity of the branch site element to the 3′ splice site in AT-AC I introns creates a steric block that prevents commitment via the GT-AG pathway. On the other hand, the loose branch site sequence requirement for the GT-AG pathway usually means that many cryptic branch sites are available for use (Padgett et al., 1985; Reed & Maniatis, 1985; Ruskin et al., 1985); moreover, because U2 is much more abundant than U12, this is unlikely to be the rate-limiting step. The most conserved element in AT-AC I introns appears to be the 5′ splice site, which is substantially different from the GT-AG 5′ splice site consensus element. It may be that its ability to be recognized by U11 and U6atac, as proposed (Hall & Padgett, 1994; Tarn & Steitz, 1996b), is much more favorable energetically than its recognition by U1 and U6, thus resulting in commitment to the AT-AC I pathway. Even when U6atac is inactivated, an AT-AC I intron fails to commit to the GT-AG pathway (Tarn & Steitz, 1996b; this study), indicating that this is not a finely tuned competition between two possible pathways.

Finally, unique protein components may play a critical role in targeting each type of intron to a particular pathway. Nothing is known at present about the protein factors that are involved in AT-AC intron splicing. Further insights into the evolution of AT-AC introns and their role in gene expression may be obtained when some of these protein components are identified.

MATERIALS AND METHODS

Construction of plasmids

A portion of the human SCN5A gene was amplified by PCR from human total genomic DNA (Promega) with primers containing restriction sites and matching exons 25 and 26 to generate a fragment comprising exon 25, intron 25, and exon 26, followed by the first 6 nt of intron 26 (Wang et al., 1996). This fragment was digested with *Hinc* II and *Xba* I and subcloned into the corresponding sites of pSP64 (Promega) to generate the pSP64-SCN5A plasmid. For construction of pSP64-SCN5AS, a different downstream PCR primer was used to amplify a fragment lacking the 6 nt of intron 26; this fragment was subcloned as a *Pst* I–*Xba* I fragment in pSP64. Likewise, plasmid pSP64-SCN4AS was constructed from a PCR fragment containing exon 2, intron 2, and exon 3 of SCN4A (George et al., 1993), subcloned into the *Hind* III and *Bam*H I sites of pSP64. Similar constructs were made for the intron 21 and flanking exons of SCN4A, with and without a portion of the 5′ splice site of intron 22. All constructs were confirmed by sequence analysis and the sequences of the splice sites were in agreement with those reported by Wang et al. (1996) and by George et al. (1993). pSP64-SCN5A and pSP64-SCN5AS were linearized with *Xba* I, and pSP64-SCN4AS with *Bam*H I, for use as templates for in vitro transcription with SP6 RNA polymerase. The transcripts contain short extensions at both ends, derived from the restriction site or vector.

Substrate preparation and in vitro splicing assays

Nuclear extract preparation and conditions for in vitro transcription and splicing of SCN4A and SCN4AS pre-mRNAs with the AT-AC I intron 2 were as described (Wu & Krainer, 1996). The conditions for splicing the SCN5A and SCN5AS pre-mRNAs with the AT-AC II intron 25 were the same, except that the incubation time was 5 h. No splicing was detectable when ATP or $MgCl_2$ was omitted. RNA was recovered and analyzed on 4.5% denaturing polyacrylamide gels, followed by autoradiography. To sequence the spliced mRNAs across the spliced junctions, each gel-purified RNA was amplified by RT-PCR with exon 25 and 26 primers (TATGAAGAGCAGCCTCAGT and CAGGGGCCGTGGGATGG). To sequence the authentic and cryptic spliced mRNAs across the spliced junctions for SCN4AS, each gel-purified RNA was amplified by RT-PCR with exon 2 and exon 3 primers (TCATCGTACTCAACAAGG and TACTCCACATTCTTGGAC). The amplified fragment, subcloned into PCR2.1 (Invitrogen), was sequenced with T7 Sequenase 2.0 (USB).

Oligonucleotide-directed RNase H cleavage and inhibition by 2′-O-methyl oligonucleotides

Oligonucleotide-directed RNase H cleavage experiments were performed as described (Wu & Krainer, 1996), using oligo-

nucleotides complementary to U1 snRNA positions 2–11, U2 snRNA positions 1–15, U4 snRNA positions 64–77, U6 snRNA positions 27–46, U12 snRNA positions 11–24 (Baserga & Steitz, 1993), U4atac snRNA positions 63–82, or U6atac snRNA positions 66–79 (Tarn & Steitz, 1996b). 2′-O-methyl oligonucleotide inhibition experiments were performed by pre-incubating the nuclear extract with 2′-O-methyl oligonucleotides under splicing conditions for 10 min, followed by the addition of the pre-mRNA and further incubation (Tarn & Steitz, 1996a). The 2′-O-methyl oligonucleotides (Oligos Etc., Inc.) were either complementary to U5 snRNA positions 34–47 (Baserga & Steitz, 1993) or a control sequence of the same length (CGUAUACCCUUGAC). All snRNA cleavage experiments were performed at least twice, with reproducible results. The inhibition experiments gave consistent results with four different AT-AC II intron constructs.

NOTE ADDED IN PROOF

Intron 1 of the chicken parvalbumin gene has non-consensus splice sites (Zhou et al., Genbank U06075) and has the typical sequence features of an AT-AC II intron (G. Latter, Q.W. & A.R.K., unpublished). U11 snRNA has recently been shown to function in recognition of an AT-AC I intron 5′ splice site in vivo (Kolossova & Padgett, 1997, *RNA 3*:227–233).

ACKNOWLEDGMENTS

We are grateful to Dr. Daniel Ramón for the unpublished *A. tubingensis* sequence, to Drs. Rachel Rempel and Joseph Nevins for the unpublished mouse E2F4 sequence, and to Dr. Richard Spritz for sharing the HPS sequences prior to publication. We thank Dr. Ramón, Dr. Bernhard Redl, and Dr. Kiyoshi Ito for valuable discussions about xylanase introns, M. Murray for comments on the manuscript, and A. Mayeda, H.-X. Liu, and other members of the laboratory for useful suggestions. This work was supported by NIH grant GM42699.

Received February 28, 1997; returned for revision March 25, 1997; revised manuscript received March 31, 1997

REFERENCES

Ares M Jr, Weiser B. 1995. Rearrangement of snRNA structure during assembly and function of the spliceosome. *Prog Nucleic Acid Res Mol Biol 50*:131–159.

Bailin T, Oh J, Feng GH, Fukai K, Spritz RA. 1997. Organization and nucleotide sequence of the human Hermansky-Pudlak syndrome (HPS) gene. *J Invest Derm*. Forthcoming.

Baserga SJ, Steitz JA. 1993. The diverse world of small ribonucleoproteins. In: Gesteland RF, Atkins JF, eds. *The RNA world*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press. pp 359–381.

Berget SM. 1995. Exon recognition in vertebrate splicing. *J Biol Chem 270*:2411–2414.

Berget SM, Robberson BL. 1986. U1, U2, and U4/U6 small nuclear ribonucleoproteins are required for in vitro splicing but not polyadenylation. *Cell 46*:691–696.

Black DL. 1995. Finding splice sites within a wilderness of RNA. *RNA 1*:763–771.

Black DL, Chabot B, Steitz JA. 1985. U2 as well as U1 small nuclear ribonucleoproteins are involved in premessenger RNA splicing. *Cell 42*:737–750.

Black DL, Pinto AL. 1989. U5 small nuclear ribonucleoprotein: RNA structure analysis and ATP-dependent interaction with U4/U6. *Mol Cell Biol 9*:3350–3359.

Black DL, Steitz JA. 1986. Pre-mRNA splicing in vitro requires intact U4/U6 small nuclear ribonucleoprotein. *Cell 46*:697–704.

Blencowe BJ, Sproat BS, Ryder U, Barabino S, Lamond AI. 1989. Antisense probing of the human U4/U6 snRNP with biotinylated 2′-OMe RNA oligonucleotides. *Cell 59*:531–559.

Chanfreau G, Legrain P, Dujon B, Jacquier A. 1994. Interaction between the first and last nucleotides of pre-mRNA introns is a determinant of 3′ splice site selection in *S. cerevisiae*. *Nucleic Acids Res 22*:1981–1987.

Crispino JD, Blencowe BJ, Sharp PA. 1994. Complementation by SR proteins of pre-mRNA splicing reactions depleted of U1 snRNP. *Science 265*:1866–1869.

Crispino JD, Sharp PA. 1995. A U6 snRNA:pre-mRNA interaction can be rate-limiting for U1-independent splicing. *Genes & Dev 9*:2314–2323.

Deirdre A, Scadden J, Smith CW. 1995. Interactions between the terminal bases of mammalian introns are retained in inosine-containing pre-mRNAs. *EMBO J 14*:3236–3246.

Feng GH, Bailin T, Oh J, Spritz RA. 1997. Mouse pale ear (ep) is homologous to human Hermansky–Pudlak syndrome gene. *Hum Mol Genet*. Forthcoming.

Fu XD. 1995. The superfamily of arginine/serine-rich splicing factors. *RNA 1*:663–680.

George AL Jr, Iyer GS, Kleinfield R, Kallen RG, Barchi RL. 1993. Genomic organization of the human skeletal muscle sodium channel gene. *Genomics 15*:598–606.

Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, Galibert F, Hoheisel JD, Jacq C, Johnston M, Louis EJ, Mewes HW, Murakami Y, Philippsen P, Tettelin H, Oliver SG. 1996. Life with 6000 genes. *Science 274*:546, 563–567.

Goldin AL. 1995. Voltage-gated sodium channels. In: North RA, ed. *Ligand- and voltage-gated ion channels*. Boca Raton, Florida: CRC Press, Inc. pp 73–111.

Gurr SJ, Unkles SE, Kinghorn JR. 1987. The structure and organization of nuclear genes of filamentous fungi. In: Kinghorn JR, ed. *Gene structure in eukaryotic microbes*. Oxford: IRL Press. pp 93–139.

Guthrie C. 1991. Messenger RNA splicing in yeast: Clues to why the spliceosome is a ribonucleoprotein. *Science 253*:157–163.

Haas H, Friedlin E, Stoffler G, Redl B. 1993. Cloning and structural organization of a xylanase-encoding gene from *Penicillium chrysogenum*. *Gene 126*:237–242.

Hall SL, Padgett RA. 1994. Conserved sequences in a class of rare eukaryotic nuclear introns with non-consensus splice sites. *J Mol Biol 239*:357–365.

Hall SL, Padgett RA. 1996. Requirement of U12 snRNA for in vivo splicing of a minor class of eukaryotic nuclear pre-mRNA introns. *Science 271*:1716–1718.

Hille B. 1992. *Ionic channels of excitable membranes, 2nd ed*. Sunderland, Massachusetts: Sinauer Associates Inc.

Hoffman BE, Grabowski PJ. 1992. U1 snRNP targets an essential splicing factor, U2AF65, to the 3′ splice site by a network of interactions spanning the exon. *Genes & Dev 6*:2554–2268.

Hoffman EP. 1995. Voltage-gated ion channelopathies: Inherited disorders caused by abnormal sodium, chloride, and calcium regulation in skeletal muscle. *Annu Rev Med 46*:431–441.

Hogan K, Gregg RG, Powers PA. 1996. The structure of the gene encoding the human skeletal muscle α1 subunit of the dihydropyridine-sensitive L-type calcium channel (CACNL1A3). *Genomics 31*:392–394.

Hwang DY, Cohen JB. 1996. U1 snRNA promotes the selection of nearby 5′ splice sites by U6 snRNA in mammalian cells. *Genes & Dev 10*:338–350.

Inoue H, Hayase Y, Iwai S, Ohtsuka E. 1987. Sequence-dependent hydrolysis of RNA using modified oligonucleotide splints and RNase H. *FEBS Lett 215*:327–330.

Ito K, Iwashita K, Iwano K. 1992. Cloning and sequencing of the xynC gene encoding xylanase C of *Aspergillus kawachii*. *Biosci Biotech Biochem 56*:1338–1340.

Jackson IJ. 1991. A reappraisal of non-consensus mRNA splice sites. *Nucleic Acids Res 19*:3795–3798.

Jenkins RN, Osborne-Lawrence SL, Sinclair AK, Eddy RL Jr, Byers MG, Shows TB, Duby AD. 1990. Structure and chromosomal location of the human gene encoding cartilage matrix protein. *J Biol Chem 265*:19624–19631.

Kandels-Lewis S, Séraphin B. 1993. Role of U6 snRNA in 5′ splice site selection. *Science 262*:2035–2039.

Kiss I, Deak F, Holloway RG Jr, Delius H, Mebust KA, Frimberger E, Argraves WS, Tsonis PA, Winterbottom N, Goetinck PF. 1989. Structure of the gene for cartilage matrix protein, a modular protein of the extracellular matrix. Exon/intron organization, unusual splice sites, and relation to alpha chains of beta 2 integrins, von Willebrand factor, complement factors B and C2, and epidermal growth factor. *J Biol Chem 264*:8126–8134.

Koester J. 1991. Voltage-gated ion channels and generation of the action potential. In: Kandel ER, Schwartz JH, Jessell TM, eds. *Principles of neural science*. New York: Elsevier Science Publishing Co. pp 104–118.

Kohrman DC, Harris JB, Meisler MH. 1996. Mutation detection in the med and medJ alleles of the sodium channel Scn8a. Unusual splicing due to a minor class AT-AC intron. *J Biol Chem 271*:17576–17581.

Kohtz JD, Jamison SF, Will CL, Zuo P, Lührmann R, Garcia-Blanco MA, Manley JL. 1994. Protein–protein interactions and 5′-splice-site recognition in mammalian mRNA precursors. *Nature 368*:119–124.

Krainer AR, Maniatis T. 1985. Multiple factors including the small nuclear ribonucleoproteins U1 and U2 are necessary for pre-mRNA splicing in vitro. *Cell 42*:725–736.

Krämer A. 1996. The structure and function of proteins involved in mammalian pre-mRNA splicing. *Annu Rev Biochem 65*:367–409.

Krämer A, Keller W, Appel B, Lührmann R. 1984. The 5′ terminus of the RNA moiety of U1 small nuclear ribonucleoprotein particles is required for the splicing of messenger RNA precursors. *Cell 38*:299–307.

Krawczak M, Reiss J, Cooper DN. 1992. The mutational spectrum of single base-pair substitutions in mRNA splice junctions of human genes: Causes and consequences. *Hum Genet 90*:41–54.

Kreivi JP, Lamond AI. 1996. RNA splicing: Unexpected spliceosome diversity. *Curr Biol 6*:802–805.

Kuo HC, Nasim FH, Grabowski PJ. 1991. Control of alternative splicing by the differential binding of U1 small nuclear ribonucleoprotein particle. *Science 251*:1045–1050.

Lafrenière RG, Rochefort DL, Kibar Z, Fon EA, Han FY, Cochius J, Kang X, Baird S, Korneluk RG, Andermann E, Rommens JM, Rouleau GA. 1996. Isolation and characterization of GT335, a novel human gene conserved in *Escherichia coli* and mapping to 21q22.3. *Genomics 38*:264–272.

Larson RG, Henning D, Haidar MA, Jhiang S, Lin WL, Zhang WW, Busch H. 1990. Genomic structure of the human proliferating cell nucleolar protein P120. *Cancer Commun 2*:63–71.

Lesser CF, Guthrie C. 1993. Mutations in U6 snRNA that alter splice site specificity: Implications for the active site. *Science 262*:1982–1988.

Liu K, Niu L, Linton JP, Crouse GF. 1994. Characterization of the mouse Rep-3 gene: Sequence similarities to bacterial and yeast mismatch-repair proteins. *Gene 147*:169–177.

Ludwig DL, Mudgett JS, Park MS, Perez-Castro AV, MacInnes MA. 1996. Molecular cloning and structural analysis of the functional mouse genomic XPG gene. *Mamm Genome 7*:644–649.

Luukkonen BGM, Séraphin B. 1997. The roles of branchpoint-3′ splice site spacing and interaction between intron terminal nucleotides in 3′ splice site selection in *Saccharomyces cerevisiae*. *EMBO J 16*:779–792.

MacCabe AP, Fernandez-Espinar MT, de Graaff LH, Visser J, Ramón D. 1996. Identification, isolation and sequence of the *Aspergillus nidulans xln* C gene encoding the 34-kDa xylanase. *Gene 175*:29–33.

Madhani HD, Guthrie C. 1994. Dynamic RNA–RNA interactions in the spliceosome. *Annu Rev Genet 28*:1–26.

Manley JL, Tacke R. 1996. SR proteins and splicing control. *Genes & Dev 10*:1569–1579.

Mattaj IW, Habets WJ, van Venrooij WJ. 1986. Monospecific antibodies reveal details of U2 snRNP structure and interaction between U1 and U2 snRNPs. *EMBO J 5*:997–1002.

McClatchey AI, Lin CS, Wang J, Hoffman EP, Rojas C, Gusella JF. 1992. The genomic structure of the human skeletal muscle sodium channel gene. *Hum Mol Genet 1*:521–527.

Montzka KA, Steitz JA. 1988. Additional low-abundance human small nuclear ribonucleoproteins: U11, U12, etc. *Proc Natl Acad Sci USA 85*:8885–8889.

Moore JM, Query CC, Sharp PA. 1993. Splicing of precursors to messenger RNAs by the spliceosome. In: Gesteland RF, Atkins JF, eds. *The RNA world*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press. pp 303–357.

Mount SM. 1996. AT-AC introns: An attack on dogma. *Science 271*:1690–1693.

Mount SM, Burks C, Hertz G, Stormo GD, White O, Fields C. 1992. Splicing signals in *Drosophila*: Intron size, information content, and consensus sequences. *Nucleic Acids Res 20*:4255–4262.

Nakai K, Sakamoto H. 1994. Construction of a novel database containing aberrant splicing mutations of mammalian genes. *Gene 141*:171–177.

Neuman E, Sellers WR, McNeil JA, Lawrence JB, Kaelin WG Jr. 1996. Structure and partial genomic sequence of the human E2F1 gene. *Gene 173*:163–169.

Newman A. 1994. Small nuclear RNAs and pre-mRNA splicing. *Curr Opin Cell Biol 6*:360–367.

Newman AJ, Norman C. 1992. U5 snRNA interacts with exon sequences at 5′ and 3′ splice sites. *Cell 68*:743–54.

Nilsen TW. 1994. RNA–RNA interactions in the spliceosome: Unraveling the ties that bind. *Cell 78*:1–4.

Nilsen TW. 1996. A parallel spliceosome. *Science 273*:1813.

Ohshima T, Nagle JW, Pant HC, Joshi JB, Kozak CA, Brady RO, Kulkarni AB. 1995. Molecular cloning and chromosomal mapping of the mouse cyclin-dependent kinase 5 gene. *Genomics 28*:585–588.

O'Keefe RT, Norman C, Newman AJ. 1996. The invariant U5 snRNA loop 1 sequence is dispensable for the first catalytic step of pre-mRNA splicing in yeast. *Cell 86*:679–689.

Ophoff RA, Terwindt GM, Vergouwe MN, van Eijk R, Oefner PJ, Hoffman SMG, Lamerdin JE, Mohrenweiser HW, Bulman DE, Ferrari M, Haan J, Lindhout D, van Ommen GJB, Hofker MH, Ferrari MD, Frants RR. 1996. Familial hemiplegic migraine and episodic ataxia type-2 are caused by mutations in the Ca^{2+} channel gene CACNL1A4. *Cell 87*:543–552.

Padgett RA, Konarska MM, Aebi M, Hornig H, Weissmann C, Sharp PA. 1985. Nonconsensus branch-site sequences in the in vitro splicing of transcripts of mutant rabbit beta-globin genes. *Proc Natl Acad Sci USA 82*:8349–8353.

Parker R, Siliciano PG. 1993. Evidence for an essential non-Watson–Crick interaction between the first and last nucleotides of a nuclear pre-mRNA intron. *Nature 361*:660–662.

Parker R, Siliciano PG, Guthrie C. 1987. Recognition of the TAC TAAC box during mRNA splicing in yeast involves base pairing to the U2-like snRNA. *Cell 49*:229–239.

Plant KE, Hair A, Morgan GT. 1996. Genes encoding isoforms of transcription elongation factor TFIIS in *Xenopus* and the use of multiple unusual RNA processing signals. *Nucleic Acids Res 24*:3514–3521.

Plassart E, Reboul J, Rime CS, Recan D, Millasseau P, Eymard B, Pelletier J, Thomas C, Chapon F, Desnuelle C, Confavreux C, Bady B, Martin JJ, Lenoir G, Serratrice G, Fardeau M, Fondaine B. 1994. Mutations in the muscle sodium channel gene (SCN4A) in 13 French families with hyperkalemic periodic paralysis and paramyotonia congenita: Phenotype to genotype correlations and demonstration of the predominance of two mutations. *Eur J Hum Genet 2*:110–124.

Reed R. 1996. Initial splice-site recognition and pairing during pre-mRNA splicing. *Curr Opin Genet Dev 6*:215–220.

Reed R, Maniatis T. 1985. Intron sequences involved in lariat formation during pre-mRNA splicing. *Cell 41*:95–105.

Robberson BL, Cote GJ, Berget SM. 1990. Exon definition may facilitate splice site selection in RNAs with multiple exons. *Mol Cell Biol 10*:84–94.

Ruskin B, Greene JM, Green MR. 1985. Cryptic branch point activation allows accurate in vitro splicing of human beta-globin intron mutants. *Cell 41*:833–844.

Sato S, Hotta Y, Tabata S. 1995. Structural analysis of a recA-like gene in the genome of *Arabidopsis thaliana*. *DNA Res 2*:89–93.

Seiwert SD, Steitz JA. 1993. Uncoupling two functions of the U1 small nuclear ribonucleoprotein particle during in vitro splicing. *Mol Cell Biol 13*:3135–3145.

Soldatov NM. 1994. Genomic structure of human L-type Ca^{2+} channel. *Genomics* 22:77–87.

Sontheimer EJ, Steitz JA. 1992. Three novel functional variants of human U5 small nuclear RNA. *Mol Cell Biol* 12:734–746.

Sontheimer EJ, Steitz JA. 1993. The U5 and U6 small nuclear RNAs as active site components of the spliceosome. *Science* 262:1989–1996.

Sun JS, Manley JL. 1995. A novel U2-U6 snRNA structure is necessary for mammalian mRNA splicing. *Genes & Dev* 9:843–854.

Talerico M, Berget SM. 1994. Intron definition in splicing of small *Drosophila* introns. *Mol Cell Biol* 14:3434–3445.

Tarn WY, Steitz JA. 1994. SR proteins can compensate for the loss of U1 snRNP functions in vitro. *Genes & Dev* 8:2704–2717.

Tarn WY, Steitz JA. 1996a. A novel spliceosome containing U11, U12, and U5 snRNPs excises a minor class (AT-AC) intron in vitro. *Cell* 84:801–811.

Tarn WY, Steitz JA. 1996b. Highly diverged U4 and U6 small nuclear RNAs required for splicing rare AT-AC introns. *Science* 273:1824–1832.

Valcarcel J, Gaur RK, Singh R, Green MR. 1996. Interaction of U2AF65 RS region with pre-mRNA of branch point and promotion base pairing with U2 snRNA. *Science* 273:1706–1709.

Wang Q, Li Z, Shen J, Keating MT. 1996. Genomic organization of the human SCN5A gene encoding the cardiac sodium channel. *Genomics* 34:9–16.

Wang Q, Shen J, Splawski I, Atkinson D, Li Z, Robinson JL, Moss AJ, Towbin JA, Keating MT. 1995. SCN5A mutations associated with an inherited cardiac arrhythmia, long QT syndrome. *Cell* 80:805–811.

Wassarman KM, Steitz JA. 1992. The low-abundance U11 and U12 small nuclear ribonucleoproteins (snRNPs) interact to form a two-snRNP complex. *Mol Cell Biol* 12:1276–1285.

Watanabe A, Ikejima M, Suzuki N, Shimada T. 1996. Genomic organization and expression of the human MSH3 gene. *Genomics* 31:311–318.

Wu HJ, Gaubier-Comella P, Delseny M, Grellet F, Van Montagu M, Rouzé P. 1996. Non-canonical introns are at least 10^9 years old. *Nature Genet* 14:383–384.

Wu JY, Maniatis T. 1993. Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. *Cell* 75:1061–1070.

Wu Q, Krainer AR. 1996. U1-mediated exon definition interactions between AT-AC and GT-AG introns. *Science* 274:1005–1008.

Yamada Y, Masuda K, Li Q, Ihara Y, Kubota A, Miura T, Nakamura K, Fujii Y, Seino S, Seino Y. 1995. The structures of the human calcium channel alpha 1 subunit (CACNL1A2) and beta subunit (CACNLB3) genes. *Genomics* 27:312–319.